

TARTU ÜLIKOOL

LOODUS- JA TEHNOLOOGIA TEADUSKOND

Füüsika Instituut

Arvutitehnika eriala

Vladimir Visbek

**RAKENDUS MEILIPOSTKASTI ANDMETE  
VISUALISEERIMISEKS**

Bakalaureusetöö (12 EAP)

Juhendaja:

Tõnu Tamme, MSc

Tartu 2015

## Resümee

Viimase kahekümne aasta jooksul on inimeste omavaheline suhtlemine väga muutunud. Reaalne kommunikatsioon asendub üha rohkem virtuaalsega. Tavaliste paberkirjade kirjutamine ja saatmine on ununemas ning sellesse hakati suhtuma, kui vanasse traditsiooni. Elektronkirjade ehk meilide saatmine on muutunud nii populaarseks, et ilma oma e-postita ehk kasutajakontota postiserveris on meie ühiskonnas raske hakkama saada.

Meilide saatmiseks ja vastuvõtmiseks on vajalik postiprogramm, mis võib olla veebipõhine, mida nimetatakse veebimeiliks, arvutiprogrammina kasutaja arvutis või nutiseadmel. Postiprogramme nimetatakse ka meiliklientideks või meiliprogrammideks.

Bakalaureusetöö eesmärgiks oli leida, millised on tänapäeva meiliprogrammide võimalused kasutaja postkasti visualiseerimiseks ning kirjutada valmis töötav rakendus, mis suudaks anda kasutajale graafikutena tagasisidet tema meilielu kohta.

Töös uuritud kõige kasutatavamal meilikliendil Gmailil puudub võimalus postkasti andmete põhjalikumaks analüüsimiseks ja selle põhjal graafikute ja statistika koostamiseks. Seda on võimalik teha ainult kasutades selleks spetsiaalselt programmeeritud Gmaili laiendusi ja lisafunktsioone, nagu näiteks Immersion ja Gmail Meter.

Seega valmis antud töö raames Pythoni keeles programmeeritud meilirakendus, millega on võimalik sisse logida oma Gmaili kasutajaga, valida sobivad otsingukriteeriumid ja analüüsimise parameetrid, allalaadida nende põhjal oma postkastist kirju, kuvada need ekraanile ning luua nende kohta seitse huvitavat graafikut. Kõik nimetatud funktsionaalsused töötasid ka lõpptestimise ajal.

Testimise ja teiste meilirakendustega võrdlemise järel jõuti järeldusele, et EDA programmiga on võimalik analüüsida kasutaja Gmaili postkasti andmed ning näidata tulemust visualiseeritud graafikutena.

# Sisukord

Resümee.....	2
Jooniste loetelu.....	5
Lühendid, konstandid, mõisted.....	6
1 Sissejuhatus .....	7
2 Ülevaade probleemist .....	8
2.1 Tänapäeva meilikliendid .....	8
2.2 Gmaili ülevaade.....	10
2.3 Gmaili postkasti visualiseerimise võimalused .....	10
2.3.1 Immersion .....	11
2.3.2 Gmail Meter .....	15
3 Metoodika.....	19
3.1 Programmeerimiskeel.....	19
3.2 Meiliprotokollid .....	19
3.3 Meiliserverist kirjade saamine .....	20
3.3.1 Ühendus serveriga.....	20
3.3.2 Meilikausta valimine.....	21
3.3.3 Meilikirjade otsing.....	22
3.3.4 Meilikirjade sisu saamine .....	22
3.4 Andmete töötlemine .....	23
3.5 Graafiline kasutajaliides.....	23
3.6 Graafikute loomine.....	24
4 Tulemus .....	25
4.1 Graafiline kasutajaliides.....	25
4.1.1 Logimisaken.....	25
4.1.2 Põhiaken.....	26
4.2 Kirjade importimine .....	28
4.3 Statistika ja graafikute loomine.....	30
5 Tulemuste analüüs ja järeldused.....	31

5.1	EDA tulemused ja analüüs .....	31
5.2	Võrdlus teiste meiliklientidega .....	38
5.3	Järeldused .....	39
	Kokkuvõte .....	40
	Summary .....	41
	Tänuavaldused .....	42
	Viited .....	43
	Lisad .....	46
	Lihtlitsents .....	47

## Jooniste loetelu

Joonis 2.1: Tänapäeva populaarsemad meilikliendid (aprill 2015) .....	8
Joonis 2.2: Populaarsemad meilikliendid (september 2014) .....	9
Joonis 2.3: Kontaktide visualisatsioon Immersionis .....	12
Joonis 2.4: Meilide statistika .....	13
Joonis 2.5: Kindla kontakti ühendused võrgustikus .....	13
Joonis 2.6: Kindla kontakti kohta statistika .....	14
Joonis 2.7: Koguste statistika .....	16
Joonis 2.8: Igapäevane aktiivsus .....	17
Joonis 2.9: Sõnade arv .....	18
Joonis 3.1: Ühendus meiliserveriga ja sisselogimine .....	21
Joonis 4.1: EDA logimisaken .....	25
Joonis 4.2: EDA põhiaken .....	26
Joonis 4.3: EDA infoaken .....	27
Joonis 4.4: EDA parameetrite aken .....	29
Joonis 5.1: Kõik postkasti kirjad viimase aasta lõikes .....	31
Joonis 5.2: Kõik saadetud kirjad viimase aasta lõikes .....	32
Joonis 5.3: Kirjade arv kuude lõikes .....	33
Joonis 5.4: Kirjade arv kellaaegade lõikes .....	33
Joonis 5.5: Kõik kirjad kuude kaupa .....	34
Joonis 5.6: Top kuud saadetud ja saadud kirjade kohta .....	34
Joonis 5.7: Kõik kirjad nädalapäevade kaupa .....	35
Joonis 5.8: Top nädalapäevad saadetud ja saadud kirjade kohta .....	35
Joonis 5.9: Sõnade arvude tulpdiaagramm .....	36
Joonis 5.10: Top 10 meilikasutajat, kellele autor kirju saatis .....	37
Joonis 5.11: Top 10 meilikasutajat, kes on autorile kirju saatnud .....	37

## Lühendid, konstandid, mõisted

**Offline** – olukord, kus arvutil pole ühenduses internetiga. Selle vastand on *online*.

**To** – väli meilikirjas, kuhu sisestatakse Saaja meiliaadress.

**From** – väli meilikirjas, kuhu sisestatakse Saatja meiliaadress.

**Cc** – ingl *Carbon Copy* on väli meilikirjas, kuhu sisestatakse kirja koopia Saaja. [1]

**Timestamp** – ajatempel on aeg, millal meilikiri oli saadetud või vastuvõetud. [2]

**Socket** – hosti sokkel on sidekanali otspunktina toimiv IP-aadressist ja pordinumbrist koosnev identifikaator. [3]

**Inbox** – sisendkast ehk koht, kus hoitakse saabunud sõnumeid.

**False** – inglise keelest *Väär*.

**True** – inglise keelest *Tõene*.

**GUI** – ingl *Graphical User Interface* ehk Graafiline kasutajaliides. [4]

**EDA** –ingl *Email Data Analysis* ehk Meiliandmete analüüs.

# 1 Sissejuhatus

Viimase kahekümne aasta jooksul on inimeste omavaheline suhtlemine väga muutunud. Reaalne kommunikatsioon asendub üha rohkem virtuaalsega. Tänapäeva tehnoloogia võimaldab suhelda inimestega mitte ainult kodust väljumata, vaid ka diivanilt tõusmata. Selleks on näiteks mobiiltelefonid, meilikliendid, Facebook, Skype jne. Tavaliste paberkirjade kirjutamine ja saatmine on ununemas ning sellesse hakati suhtuma, kui vanasse traditsiooni. Elektronkirjade ehk meilide saatmine on muutunud nii populaarseks, et ilma oma e-postita ehk kasutajakontota postiserveris on meie ühiskonnas raske hakkama saada. Paljudes elusituatsioonides, nagu tööl või koolis, on see muutunud asendamatuks vahendiks. Nüüd on täiesti tavaline, kui inimene saab päevas 5 kuni 10 kirja.

Meilide saatmiseks ja vastuvõtmiseks on vajalik postiprogramm, mis võib olla veebipõhine, mida nimetatakse veebimeiliks, arvutiprogrammina kasutaja arvutis või nutiseadmel. Postiprogramme nimetatakse ka meiliklientideks või meiliprogrammideks. Viimastel aastatel on kõige populaarsemateks meiliklientideks saanud nutiseadmete (nutitelefonid, tahvelarvutid jne) meilikliendid. [5]

Päevast päeva, meilivahetuse käigus, koguneb postkastidesse suur kogus andmeid, mida on võimalik analüüsida, luua graafikuid ja statistikat ning tulemusena saada kasulikku informatsiooni ja tagasisidet oma meilivahetuse kohta.

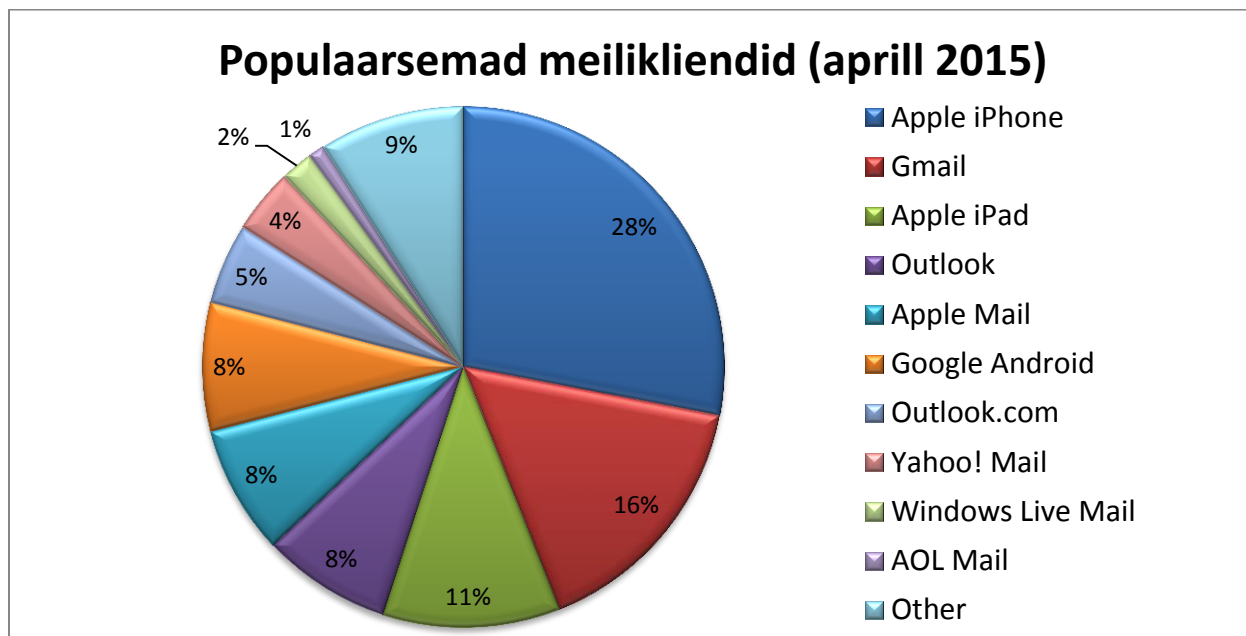
Bakalaureusetöö eesmärgiks on leida, millised on tänapäeva meiliprogrammide võimalused kasutaja postkasti visualiseerimiseks ning kirjutada valmis töötav rakendus, mis suudaks anda kasutajale graafikutena tagasisidet tema meilielu kohta.

## 2 Ülevaade probleemist

Meili teel suhtlemise populaarsuse tõttu on loodud mitmed võimalusi meilikirjade saatmiseks ja lugemiseks. Seega püütakse antud peatükis leida populaarseimat selleks kasutatavat meiliklienti ning uuritakse, kuidas ta suudab postkasti andmeid visualiseerida.

### 2.1 Tänapäeva meilikliendid

Joonise 2.1 pirukadiagrammil on edetabel populaarsematest meiliklientidest, mis oli koostatud Litmus Email Analytics poolt kogutud andmetest aprillis 2015. Antud meiliklientide seas on ka veebimeilid ja nutiseadmete meilirakendused. Andmed saadi kätte 1,01 miljardi meilikirja avamisest. [6]

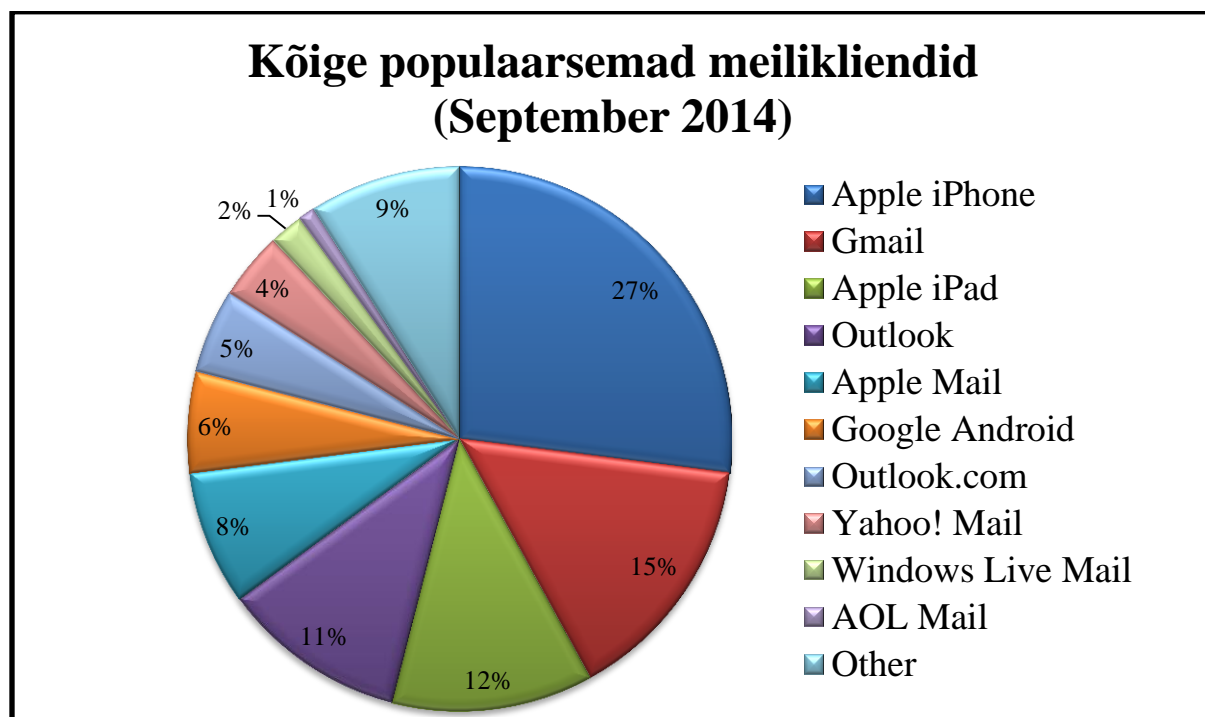


Joonis 2.1: Tänapäeva populaarsemad meilikliendid (aprill 2015)



Nagu on diagrammilt näha, kõige rohkem meilikirju, umbes üks kolmandik, on avatud Apple iPhone seadmel. Teisel kohal on 16 protsendiga Gmail veebimeil. Kolmandal ja neljandal on umbes võrdse avamiste arvuga Apple iPad ja Outlook, vastavalt 11% ja 8%. Ülejäänud meiliklientides on avatud alla 10% meilikirju: Apple Mail 8%, Google Android 8%, Outlook.com 5%, Yahoo! Mail 4%, Windows Live Mail 2% ja AOL Mail 1%. Lisaks on ka muid meiliprogramme, millega on avatud ligi 9% meilidest. [6]

Võrdlemiseks on joonisel 2.2 toodud välja ka samasugune edetabel andmetest, mis olid kogutud 2014. aasta septembris. [6]



Joonis 2.2: Populaarsemad meilikliendid (september 2014)

Antud andmete järgi võib teha järelduse, et nutiseadmete meilikliendid on väga populaarsed. Siiski töölaua meiliprogramme ja veebimeile kasutatakse rohkem (umbes 53%). Samuti on statistikast näha, et töölaua meilirakendusi kasutatakse üldiselt vähe ja nende protsent langeb.

Seega edaspidi keskendutakse kõige populaarsemale veebimeilile Gmailile ja selle laiendustele ehk lisafunktsioonidele.

## 2.2 Gmaili ülevaade

Gmail on Google korporatsiooni tasuta meiliteenuste pakkuja. [7]

See veebimeil tagab põhimõtteliselt lõpmatut tasuta online mäluruumi kasutaja sõnumite hoidmiseks, mida võibki pidada selle peamiseks populaarsuse põhjuseks. Gmail tekkel pakkusid teised veebimeilid palju väiksemat mäluruumi. Samuti on Gmail lihtsa kasutajaliidesega: kirjade kirjutamine ja leidmine ei nõua palju pingutust. Gmail Gears võimaldab lugeda ja kirjutada meilikirju offline'is. Lisaks, saab muuta oma postkasti tausta (värvilised teemad, HD-teemad jne) ning kohandada selle elementide paiknemist vastavalt oma nõuetele. [7]

Suureks plussiks on see, kuidas Google on ühendanud kõik oma teenused. Olles sisselogitud Gmaili, saab ühe klikiga minna näiteks Google+, Youtube'i, Google Mapsi ühe kasutajakonto all. [7]

## 2.3 Gmaili postkasti visualiseerimise võimalused

Algselt on Gmaili meilipostkasti kasutajaliides erinenud teistest veebimeilidest. See keskendus otsingumootorile ja vestluste hargtöötlusele (sõnumite grupeerimine nende vastustega) [8], mida praeguseks on teised konkurendid ka kasutusele võtnud. Samuti kasutab Gmail vahekaarte, kuhu saab kirju jaotada:

- **Peamine** – isiklikud vestlused ja kirjad, mida ei kuvata teistel vahekaartidel;
- **Suhtlusvõrgustikud** – kirjad suhtlusvõrgustikest, meediajagamise saitidel, veebikohtamisteenustest ja muudelt sotsiaalveebisaitidelt;
- **Pakkumised** – erinevad pakkumised ja muud reklaamimeilid;

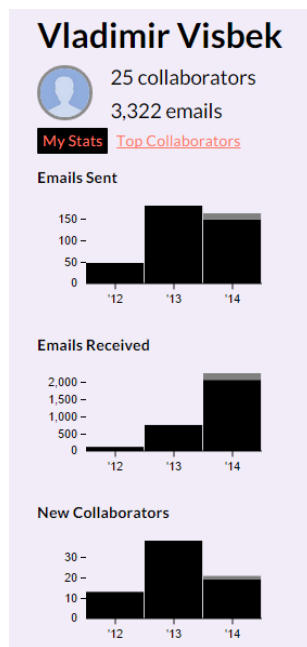
- **Uuendused** – isiklikud ja automaatselt loodud uuendused, sh kinnitused, kviitungid, arved ja väljavõtted;
- **Foorumid** – kirjad veebigruppidelt, foorumitest ja meililistidest. [7,9]

Kirju vahekaartides näidatakse lihtsalt aja järgi järjestatud listina ning puudub võimalus muud moodi andmeid visualiseerida ja nende põhjal statistikat luua. Seetõttu on loodud mitmed veebirakendused, –laiendused ja programmid, mis suudavad analüüsida Gmaili meilikasti andmeid ning huvitavalt neid presenteerida. [7]

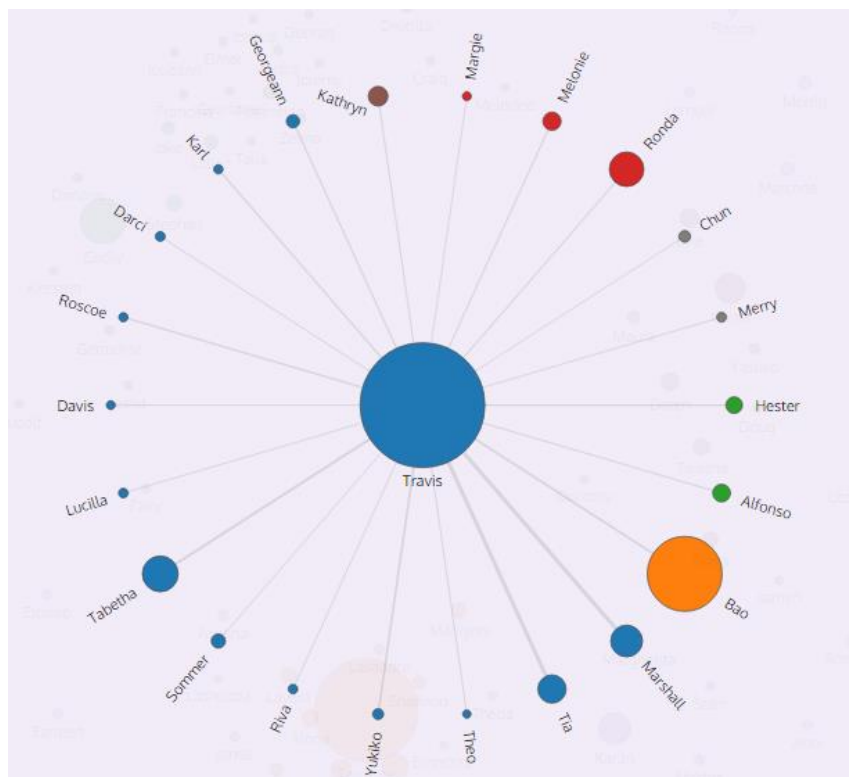
### 2.3.1 Immersion

2013. Aasta suvel kuulutas Massachusettsi Tehnoloogia Instituudi Meedia labori meeskond välja uut tööriista, mis kasutab Gmaili meta-andmeid visuaalse meilivahetuse võrgustiku loomiseks. Antud veebirakendus võtab kõikidelt meilidelt ainult *To*, *From*, *CC* ja *Timestamp* andmed ning loob nende abil kasutaja elu ja suhete võrgustikku [10]. Jooniselt 2.3 on näha, et programm kujutab kontakte erinevat värvi ringidena ning nende suurus erineb vastavalt sellele, kui palju on kasutaja kindla inimesega kirju vahetanud. See esitab suhtlemisandmeid kõikide sõpradega, tuttavatega, pereliikmetega ja kaastöötajatega alates meiliaadressi loomisest. Samuti grupeerib programm automaatselt kontakte ning ühendab need omavahel joontega. Selline ühendus tekib, kui kirju saadetakse korraga mitmele kasutajale. [11]

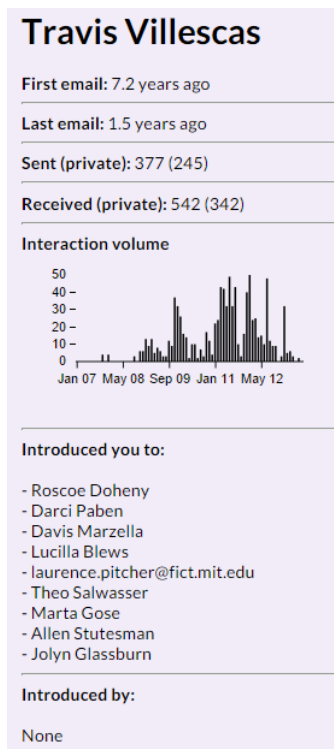




Joonis 2.4: Meilide statistika



Joonis 2.5: Kindla kontakti ühendused võrgustikus



Joonis 2.6: Kindla kontakti kohta statistika

Vajutades kindlale ringile, näidatakse lähemat statistikat antud kontaktiga, mida saab näha joonistelt 2.5 ja 2.6: esimese ja viimase meilikirja kuupäev, saadetud ja vastuvõetud kirjade hulk, tulpdiagramm kõikidest saadetud kirjadest ja suhted teiste kontaktidega. [11]

Samuti saab antud statistikat piirata kindla ajavahemikuga, valides kas üks päev, üks kuu või kogu meiliaadressi eksisteerimise periood. [11]

Kindlasti sellisel meetodil on ka oma limiidid. Kasutajal võib olla mitu meiliaadressit, mille tõttu ta võib neid kõiki kasutada võrdväärselt ja ei teki üldist meilivahetuse pilti. Sarnane probleem võib tekkida ka siis, kui kirju regulaarselt kustutatakse. Mõnel tuttavale ei pruugi olla oma meili või kõigiga ei suheldagi meili teel. Immersion ei mõõda meilisuhete kvaliteeti, vaid kogust. [10]

Immersion kasutab kasutajate Google veebilehega autoriseerimiseks tehnoloogiat Oauth. Meiliaadress ja parool, mida sisestatakse programmi kasutamisel, ei ole Immersionile nähtavad,

kuna logimislehte kontrollib Google. Kui Google kinnitab sisestatud andmed, siis ta teatab sellest Immersion programmile ning kasutajale antakse võimalus lubada või keelata programmi ligipääs meiliandmetele. [11]

Väljalogimisel pakub Immersion kustutada kogutud meta-andmeid. Kustutamisel andmeid ei märgistata, kui *kustutatud* või ei edastata muule serverile, vaid koheselt kustutakse antud serverilt. Samuti ei jagata isiklikke andmeid kolmanda osapoolega. Samas, kui meta-andmeid ei kustutata, siis järgmisel sisselogimisel ei pea Immersion uuesti neid analüüsima ning võrgustiku laadimine on kiirem. [11]

Selleks, et MIT Media Lab server üle ei koormaks, on Immersion programmil ülemine limit 300 000 kirja kasutaja kohta. Selline number valiti testimisel ja kasutajate tagasisidel. [11]

Lisaks, et mingi kindel kontakt oleks visualiseerimisel nähtav peab kasutajaga nende vahel olema vähemalt 3 saadetud ja vastuvõetud meilikirja. Selline lähenemine filtreerib visualisatsioonist välja meililistid, sotsiaalsete võrgustike teated, spämmid jne. Vaikimisi näitab võrgustik top 100 kontakti, aga liigutades lehe vasakul pool olevaid liugureid on neid võimalik rohkem lisada. [11]

Immersion programmiga võib luua visualisatsiooni Gmail'i, Yahoo ja MS Exchange meilipostkastidest. [11]

Immersioni töövõimekuse näitamiseks kasutatud pildid on loodud autori isiklikust postkastist ja programmi enda demost, mida on võimalik laadida selle kodulehel. [11]

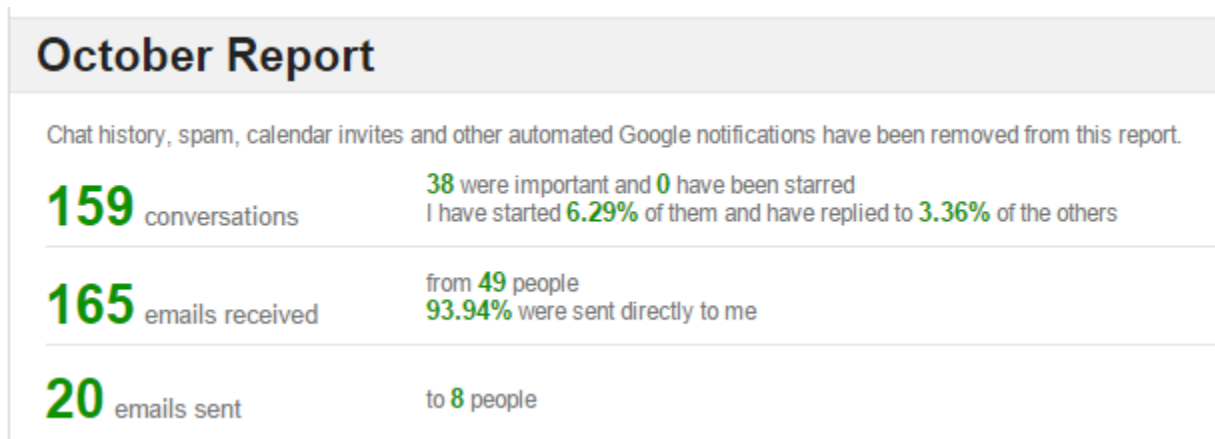
### **2.3.2 Gmail Meter**

Gmail Meter on Google Chrome brauseri rakendus ja Gmaili laiendus. Selle installimine ja ligipääsu andmine meilipostkastile võimaldab saada huvitavat statistikat Gmail andmete kohta. Programmeeritud Apps Scripti keeles, käivitub see iga kuu esimesel päeval ja saadab eelmise kuu andmete põhjal tehtud statistika meilikirjana kasutaja postkasti. Statistika andmed

võimaldavad analüüsida meilipostkasti kasutamise harjumusi ja aktiivsust. Spämmi, kalendri ja Google teateid aruanne ei sisalda. [12,13]

Järgnevalt loetletakse täpsemalt mõned näited, mida postkasti analüüs tagastab.

**Koguste statistika** – tähtsate ja lemmikute kirjade kogus, saadetud ja vastuvõetud kirjade maht, mitmele inimesele saadetud, mitmelt inimeselt saadud jne (joonis 2.7).



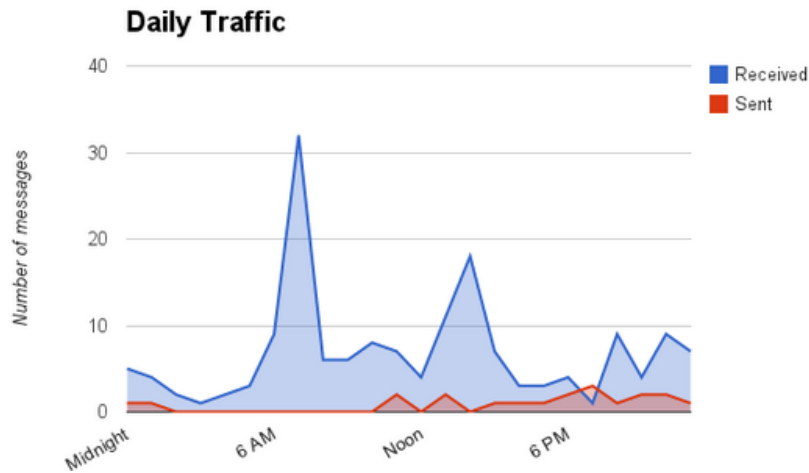
Joonis 2.7: Koguste statistika

**Top statistika** – nimekiri inimestest, kellele on meilikirju kõige rohkem saadetud, kellelt vastuvõetud ja mis teemal. [13]

**Igapäevane aktiivsus** näitab ajadiagrammina millal kasutaja on kirju saatnud ja vastuvõtnud päeva jooksul kõige rohkem. Näiteks jooniselt 2.8 on näha, et kõige rohkem kirju on saadud vara hommikul, kuna siis saadetakse automaatselt teateid ja reklaame erinevatelt veebilehtedelt. [13]

Sarnaselt selle diagrammiga on aruandes ka **iganädalane** ja **igakuine statistika**. [13]



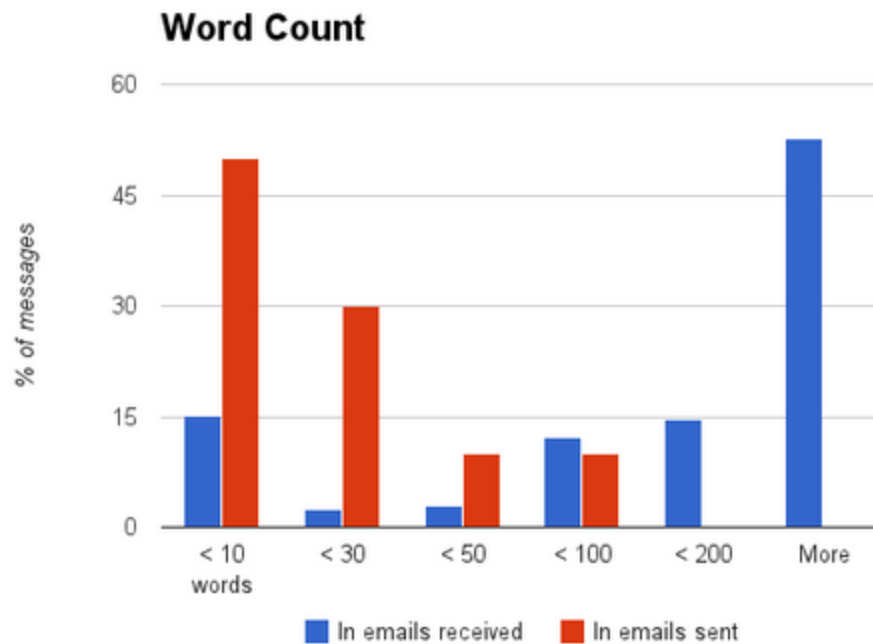


Joonis 2.8: Igapäevane aktiivsus

**Meilikategooriad** – pirukadiagramm, mis näitab kuidas kasutaja oma postkasti kasutab. Kirjad on jagatud gruppidesse: *märgistatud*, *prügikast*, *arhiveeritud* ja *postkast*. [13]

**Aeg enne vastust** – tulpdiagramm, mis näitab, kui kiiresti vastab kasutaja vastuvõetud kirjadele ja kui kiiresti vastatakse talle. [13]

**Sõnade arv** – tulpdiagramm, mis näitab, kui tihti ja mis pikkusega kirju kasutaja kirjutab või saab vastu. Näiteks joonisel 2.9. on näha, et kõige rohkem saadetakse kirju pikkusega alla 10 sõna. [13]



Joonis 2.9: Sõnade arv

**Manuste statistika** – tulpdiagramm, mis näitab, kui tihti on mingit kindlat tüüpi failid manustena kirjadega küljes. [13]

## 3 Metoodika

### 3.1 Programmeerimiskeel

Kuna autoril on olnud kõige rohkem kogemusi Python programmeerimiskeelega ning sellele on olemas mooduleid peaaegu kõige jaoks, siis meilirakendus tehti just selles keeles.

### 3.2 Meiliprotokollid

Kõik vastuvõetud ja saadetud meilid asuvad kasutaja meilipostkastis meiliserveris. Elektronkirjade saamiseks saab meiliklient pääseda meilipostkasti otse või allalaadida serverist kirju kasutaja arvutisse ja hallata neid seal edasi. Kasutaja meilipostkastiga võib ühendust saada kasutades kahte protokollit: POP või IMAP. Enamus tänapäeva meiliklientidest toetavad mõlemat protokollit. [5]

POP (ingl *Post Office Protocol*) ehk postkontoriprotokoll on lihtsa lähenemisega. Meiliklient ühendub POP abil serveriga, saab kõik meilid kätte, salvestab need uute kirjadena kasutaja arvutis, vähemasti kustutab need serverist ja siis katkestab ühenduse. Tegelikult on enamus meiliklientidel võimalus jätta meilikirjade koopiad serverisse. See protokoll ei ole mugav kasutajate jaoks, kes soovivad lihtsalt pääseda samale meilipostkastile erinevatelt masinatelt. Probleemiks on POP ühesuunaline kommunikatsioon. Ühelt masinalt meilide lugemine või kustutamine ei muuda kirja staatust meiliserveris ning ühendumisel teise masinaga (serveris on meilid märgistatud kui *lugemata*) laetakse kõik meilid alla uuesti. Postkontoriprotokollist on olnud erinevaid versioone ning praegune standardversioon on POP3. [5,14]

IMAP (ingl *Internet Message Access Protocol*) ehk internetisõnumitele juurdepääsu protokoll on postkontoriprotokollist palju keerulisem. See annab kasutajale juurdepääsu meiliserveris hoitavatele meilikirjadele ja võimaluse hallata neid nii, nagu asuksid nad kasutaja arvutis. Erinevalt POP-ist on võimalik sõnumeid alla laadida osade kaupa (näiteks kiri ilma manuseta) ja

märgistada, kui *loetud*, *vastatud*, *kustutatud* jne. Tänu sellele on juurdepääs meilipostkasti erinevatest masinatest mugav. Samuti laeb meiliklient kasutaja arvutisse meilikirjade koopiaid ja nende originaalid jäävad alati serverisse. Samas IMAP on aeglasem, kui POP, kuna ühendumisel serveriga toimub peale meilide allalaadimist meilikliendi postkasti sünkroniseerimine meiliserveriga. Kui meilikast on liiga suur, siis kõikide kirjade allalaadimine ning sünkronisatsioon võtab palju aega ja ruumi. Viimane versioon sellest protkollist on IMAP4. [5,14]

Need protokollid võimaldavad ainult meilide vastuvõtmist. Sõnumite edastamiseks kasutavad meilikliendid SMTP prokolli (ingl *Simple mail transfer protocol*) ehk lihtsat meiliedastusprotokolli. [5]

Võttes arvesse kõik nimetatut, valiti programmi kirjutamisel meilikirjade allalaadimiseks IMAP protokoll.

### **3.3 Meiliserverist kirjade saamine**

Gmail on antud hetkel kõige populaarsem veebimeil, seega selle töö raames valmiv meiliklient on suunatud Gmaili kasutajatele.

Pythoni keeles meilide saamiseks Gmailist on mitu võimalust. Üheks võimaluseks on ka IMAP, aga serveriga suhtlemisel on mitu mehhanismi. Esimene, *IMAP4*, kasutab puhtaid socketeid; *IMAP4\_SSL* kasutab krüpteeritud kommunikatsiooni SSL Socketite kaudu; ja *IMAP4\_stream* kasutab väliste käskude standardseid sisendeid ja väljundeid. Et programmi kasutamine oleks kasutajale turvalisem, kasutati programmis *IMAP4\_SSL*-i. [15]

#### **3.3.1 Ühendus serveriga**

Ühenduse loomisel serveriga IMAP abil on kaks sammu. Esiteks, seatakse socketi ühendus. Teiseks, toimub kasutaja autentimine serveris. [15]

```

import imaplib
import os

def open_connection(verbose=False):

    # Ühendus serveriga
    hostname = 'imap.gmail.com'
    connection = imaplib.IMAP4_SSL(hostname)

    # Kasutaja autentimine ja sisselogimine
    username = 'meilikasutaja'
    password = 'parool'
    try:
        connection.login(username, password)
        print("Ühendus õnnestus!")
    except imaplib.IMAP4.error:
        print("Ühendus ebaõnnestus!")
    return connection

c = open_connection(verbose=True)

```

Joonis 3.1: Ühendus meiliserveriga ja sisselogimine

Eelkõige toimub ühendus Gmaili serveriga (joonis 3.1), mida saab teha funktsiooniga *imaplib.IMAP4\_SSL(hostname)*, mille argumentina tuleb täpsustada serveri hostinimi. Antud juhul, kuna soovitakse ühendust Gmaili serveriga IMAP abil, tuleb argumentiks *imap.gmail.com*. [15,16,17]

Peale serveriga ühendamist, tuleb sisse logida ehk autentida kasutajat serveris. Seda saab teha kasutades *login(username, password)*, mille argumentidena tuleb täpsustada kasutajanime ja selle parooli. [15,16,17]

### 3.3.2 Meilikausta valimine

Meilipostkastis asuvad meilikirjad erinevates kaustades, näiteks: *INBOX* (vaikimisi), *All*, *Trash* jne. Nende nimetused võivad esineda erinevates keeltes, vastavalt sellele, mis keeles kasutab kasutaja oma Gmaili postkasti. Samas *INBOX* kaust on sama nimega iga keelega postkastis. Et näha, millised kaustad on antud kasutajal kasutusel saab kasutada meetodit *list()*. [15,16,17]

Meilikirjadega töötamisel tuleb alati eelkõige valida meilikaust, kus need asuvad. Ühendus on olekuline ehk kui mingi kaust on valitud, siis kõik sisestatud käsud ja funktsioonid käivad selles kaustas olevate kirjade kohta, kuni mõni muu meilikaust pole valitud. Selleks saab kasutada funktsiooni *select(kaustanimi)*, kus argumendina tuleb täpsustada kausta nime. Samuti on oluline kirjutada argument topelt jutumärkidega, näiteks `""INBOX""` või `""Saadetud kirjad""`, kuna Gmail ise nõuab sellist süntaksit ja vastasel juhul tuleb veateade. [15,16,17]

### 3.3.3 Meilikirjade otsing

Peale meilikausta valimist saab kasutada meetodit *search(charset, criterion)* selles olevate meilikirjade tunnusmärkide teada saamiseks. Esimese argumendiga saab täpsustada CHARSET ehk sümbolite kodeerimist, näiteks *utf-8* või *None*, kui ei soovita. Teise argumendiga tuleb täpsustada kriteerium, mille järgi kirju otsida. Mõned näited:

- *search(None, 'ALL')* – kõik kirjad
- *search('iso-8859-1', '(SINCE "01-Jan-2012")')* – kirjad alates 1. jaanuar 2012
- *search(None, '(FROM "Peeter" SUBJECT "test")')* – inimesel nimega Peeter ja pealkirjaga „test“.

Antud meetod tagastab meilikirjade tunnusmärkide listi, mis on serveris igale kirjale määratud. [15,16,17,18]

### 3.3.4 Meilikirjade sisu saamine

Meetodi *search()* kasutamisel saadud tunnusmärke kasutatakse leitud meilikirjade sisu või selle osade saamiseks. Selleks tuleb rakendada funktsiooni *fetch()*, millel on kaks argumenti: kirja tunnusmärk ja kirja osad, mida soovitakse saada. Kirjaosade argument on IMAP list kirja segmentide nimedega. Kui teise argumendina panna protokoll nime *RFC822*, siis saab kätte

terve kirja. See väljastab sõnastiku meile vajalike andmetega, mida tuleb veel konverteerida sõnumiks funktsiooniga *email.message\_from\_bytes*. [15,16,17]

Kuna see sõnum on ka sõnastik, siis selle osad saab kergelt kätte: *msg['From']*, *msg['Date']*, *msg['Subject']* jne. Kirja sisuga ehk kehaga on keerulisem, kuna sellisel viisil, nagu teised osad on saadud, seda kätte ei saa. Eelkõige tuleb kontrollida, kas sõnumi nö „kasulik koormus“ (ingl *payload*) ehk andmepaketi osa, mis sisaldab sõnumi sisu, on alamsõnumi objektide list või on struktureerimata. Seda saab teha funktsiooniga *is\_multipart()*. Kui väljund on *False*, siis saab sõnumilt otse sisu sõnena kätte meetodiga *get\_payload()*. Kui väljund on *True*, siis saame analoogselt sisu kätte sõnumi esimesest elemendist. [15,16,17,19,20]

### 3.4 Andmete töötlemine

Kirjade elementide saamisel väljastab programm need ekraanile ning salvestab ühekaupa ka eraldi listidesse nende edaspidiseks töötlemiseks ja analüüsimiseks. Hiljem antud listid konverteeritakse sobivale kujule, saadakse kätte vajalik info ning koostatakse graafikud ja statistika.

### 3.5 Graafiline kasutajaliides

Pythoni programmeerimiskeelel on suur hulk graafilise kasutajaliidese teeke, alates TkInterist (traditsiooniliselt pakitud Pythoniga), kuni erinevate platvormist sõltumatute ja sõltuvate lahendusteni. [21]

Kuna TkInter on faktiliselt standardne ja enim kasutatavaim GUI pakett, siis antud töös kasutati seda. Selle valikut pooldas ka see fakt, et töös vajalikud funktsionaalsused ja lahendused oli võimalik Tkinteriga lihtsalt implementeerida. Näiteks sai kasutada tavalisi nuppe, märkeruute (ingl *checkbox*), tekstivälju, menüüribasid jne. [22]

### 3.6 Graafikute loomine

Jällegi on olemas suur hulk erinevaid teke graafikute loomiseks Python keeles. Antud töö nõueteks oli võimalus luua eri liiki kujundeid: histogramm, joondiaagramm, pirukas jne, ning võimalus loodud graafikuid salvestada oma arvutisse.

Paketi valimisel tugineti sellele, et oleks võimalik luua eelnevalt loetletud kujundeid ja, et nende loomine oleks võimalikult lihtne. Seega kasutati Matplotlib andmeteeki. [23]

Probleemiks võib kujuneda see, et Matplotlib kasutab ise mitu teeki, mida tuleb iseseisvalt eraldi allalaadida. Need on moodulid: *six*, *python-dateutil*, *pyparsing* ja *numpy*. Vastasel juhul teatakse veateade kujul, et kindel teek on puudu.

Matplotlib on organiseeritud hierarhiliselt, mille tipus asub *matplotlib.pyplot* mooduli poolt tagatud nõ „olekumasina keskkond“ (ingl *state-machine environment*). Sellel tasemel kasutatakse lihtsaid funktsioone antud kujundi teljeelementide lisamiseks (jooned, pildid, tekst jne). Järgmisel tasemel on objektorienteeritud liides, milles kasutatakse funktsioone kujundite ja telgede loomiseks. [23]

Kujundi ehk figuuri loomisel on võimalik selle sisse panna ka mitu telge ehk nõ mitu graafikut ühes aknas. Samuti saab loodud kujundeid salvestada ühte pdf-tüüpi faili, või eraldi erinevate pilditüüpidenä. [23]



## 4 Tulemus

Tulemusena valmis programm nimega EDA (ingl *Email Data Analysis*), millel on graafiline kasutajaliides, võimaldab sisse logida oma Gmaili kasutajaga, võimaldab importida oma valitud kriteeriumide põhjal saabunud või saadetud kirjade seast kindlaid sõnumeid, saab luua valitud kirjade kohta statistikat ja graafikuid.

### 4.1 Graafiline kasutajaliides

#### 4.1.1 Logimisaken

Programmi käivitamisel esimesena tuleb lahti logimisaken (joonis 4,1), kus kasutaja saab meilikasutajaga sisse logida, et saada ligipääsu oma meilikasti andmetele.



Joonis 4.1: EDA logimisaken

Sisselogimiseks on aknas kasutajanime ja parooli sisestamiseks vajalikud väljad. Parooli on võimalik märkeruudu abil esitada tavalise teksti kujul või peita tärnide taha. *Näita parooli* all on sarnane nupp, mis võimaldab sisestatud kasutaja meelde jätta, et järgmisel programmi käivitamisel ei peaks kasutajanime uuesti sisse kandma.

Kõige all on veel nupud väljumiseks ja sisselogimiseks. Seda on võimalik teha ka menüüst, mis asub akna ülaservas. Menüüs on ka abi saamise võimalus.

#### 4.1.2 Põhiaken

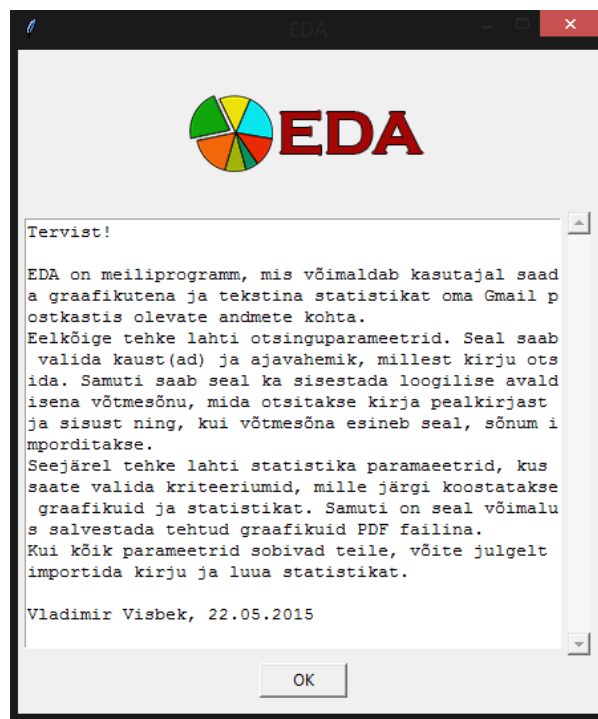
Kui sisselogimine õnnestus, tuleb lahti programmi põhiaken (joonis 4.2), kus toimub juba andmete kättesaamine ja analüüsimine.



Joonis 4.2: EDA põhiaken

Peaaegu kogu akna ruumi katab tekstiväli. Sinna väljastatakse andmed (meilikirjad) ning nende analüüsimise tulemus. Kuna andmeid võib palju tulla, on paremal pool lisatud kerimisriba, mis võimaldab reguleerida, millist andmete osa on aknas parajasti näha. Sellest veel paremal on hulk erinevaid nuppe: *Abi*, *Otsinguparameetrid*, *Statistika parameetrid*, *Impordi kirjad*, *Statistika* ja *Välju*.

Eelkõige on uuel kasutajal kõige mõttekam vajutada *Abi* nuppu, mille vajutades tuleb lahti uus väike aken. Infoaknas, nagu joonisel 4.3 näha, on samuti nagu põhiaknas suurem kerimisribaga tekstiväli, kus on lahti seletatud, kuidas EDA programmi kasutada. Sellest väljumiseks on kõige all nupp *OK*.



Joonis 4.3: EDA infoaken

Nupp nimega *Otsinguparameetrid* teeb samuti uue akna lahti, kus on võimalik valida parameetrid ja kriteeriumid, mille järgi saab meilikirju importida. Selle all on kaks nuppu

valikute salvestamiseks ja akna kinnipanemiseks. Kui aken kinni panna enne nuppu *Salvesta* vajutamist, siis parameetrite aknas tehtud muudatusi ei salvestata.

Sarnaselt eelmise nupuga töötab ka *Statistika parameetrid*. Ainult, et selle abil avanevas aknas on võimalik seadistada, kuidas kasutaja oma kirju analüüsida soovib.

Nupp *Impordi kirjad* hakkab kirju otsima ning oma töö lõpus väljastab need ekraanile. Enne importimise alustamist ilmub ekraanile väike infoaken, kus kasutajat hoiatatakse, et kirjade allalaadimine võib võtta kaua aega.

*Statistika* nupp on enne kirjade importimist deaktiveeritud, kuna ei ole veel andmeid, mida saaks analüüsida. Kohe, kui andmed on kätte saadud, muutub nupp aktiivseks ja on valmis vajutamiseks.

Viimasena on kõige all nupp *Välju*, mille abil saab põhiakna kinni panna.

Kõiki neid asju saab teha ka menüüst, mis asub akna ülaservas. Seal on võimalik ka tühjendada tekstivälja.

## **4.2 Kirjade importimine**

Enne kirjade importimist on kasutajal võimalik valida teatud parameetrid ja kriteeriumid, mille põhjal meilisõnumeid otsitakse. Need valikud saab teha otsinguparameetrite aknas (joonis 4.4), mis on lahti tehtav kas põhiakna menüüst või nupuga *Otsinguparameetrid*.

The screenshot shows a window titled 'EDA' with a search configuration interface. The main heading is 'Siin on võimalik valida parameetrid, mille järgi kirju otsitakse'. Below this, there are two sections for selecting criteria: 'Valige meilikaust(ad), millest kirju otsida:' with checkboxes for 'Postkasti kirjad' (checked) and 'Saadetud kirjad'; and 'Valige ajavahemik, millest kirju otsida:' with radio buttons for 'Viimase aasta kirjad' (selected), 'Kõik kirjad', and 'Oma vahemik'. The date range section includes 'Algne kuupäev:' (1, January, 2014) and 'Lõppkuupäev:' (1, January, 2016). At the bottom, there is a text input field for 'Sisestage võtmesõna:' and two buttons: 'Salvesta' and 'Välju'.

Joonis 4.4: EDA parameetrite aken

Eelkõige saab kasutaja valida kahe kausta vahel, kust kirju otsima hakatakse. Kuna need on tehtud märkeruutudena, saab kaustad valida kas ükshaaval või mõlemad korraga. Otsimiskriteeriumitena on võimalik määrata ajavahemik, mille seest kirju otsida – kas kõik postkastis olevad kirjad, viimase aasta kirjad (programmi kasutamise päevast) või enda poolt määratud ajavahemiku kirjad. Samuti on võimalik valida ka võtmesõnu, mida otsitakse kirja pealkirjast ja sisust ning sõnum imporditakse, kui sõna esineb vähemalt ühes nendest. Võtmesõnu saab esitada kas ükshaaval või mitme korral kasutades loogilisi tehteid *AND*, *OR* ja *NOT*, näiteks kujul *bakalaureus AND vladimir OR NOT tere AND NOT test*. Võtmesõna tekstivälja tühjaks jätmise korral ei ole see määratud ning seega otsitakse kõiki kirju.

### 4.3 Statistika ja graafikute loomine

Nupuga *Statistika parameetrid* avanevas aknas saab kasutaja seadistada kriteeriumeid, mis mõjutavad edaspidist andmete analüüsi. Kriteeriumid on esitatud märkeruutudena, mille märgistamisega näidatakse, et antud kriteerium on tarvis rakendada meilikirjade statistika loomisele. Näiteks valides märkruudud *Graafikud* ja *Histogrammid*, koosneb tagastatav statistika graafikutest, mis omakorda koosnevad histogrammidest.

Kokku on võimalik luua seitse erinevat graafikut. Mõne graafiku korral saab ka määrata, mis kujul seda esitada: pirukadiagrammina, histogrammina või joondiagrammina.

- Meilikirjade statistika ajavahemiku ja kella järgi. X-teljel on otsinguparameetrina valitud ajavahemik ning y-teljel on ööpäeva kellaajad. Graafikul on iga kiri tähistatud roheline rombina.
- Kirjade arv kellaaja lõikes. Saab esitada joondiagrammina ja/või histogrammina.
- Kirjade arv kuude lõikes. Saab esitada joondiagrammina ja/või histogrammina.
- Top 10 kuud, vaatamata valitud ajavahemikule. Saab esitada pirukadiagrammina ja/või histogrammina.
- Top 10 nädalapäeva. Saab esitada pirukadiagrammina ja/või histogrammina.
- Sõnade arvu järgi jagatud kirjad. Saab esitada histogrammina.
- Top saatjad ja saajad (kellele kasutaja saatis ja kellelt vastu võtnud).

Samuti on võimalik esitada statistikat ka tekstina. Selleks on parameetrite aknas olemas ka märkeruut *Statistika tekstina*, mille valides väljastatakse tulemus põhiprogrammi suures tekstiväljas.

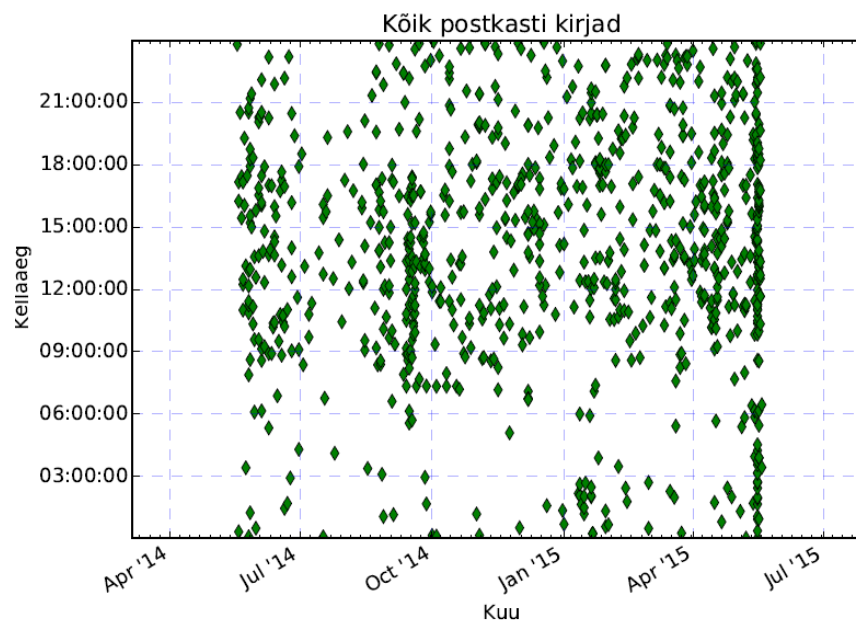
Statistika parameetrites on lisaks ka võimalus salvestada loodud graafikud PDF failina, millele saab ka oma nime anda. Selleks on akna alumises osas vastav märkeruut ja tekstiväli. Vaikimisi nimeks tuleb *EDAGraafikud*, kui tekstiväli oli tühjaks jäetud.

## 5 Tulemuste analüüs ja järeldused

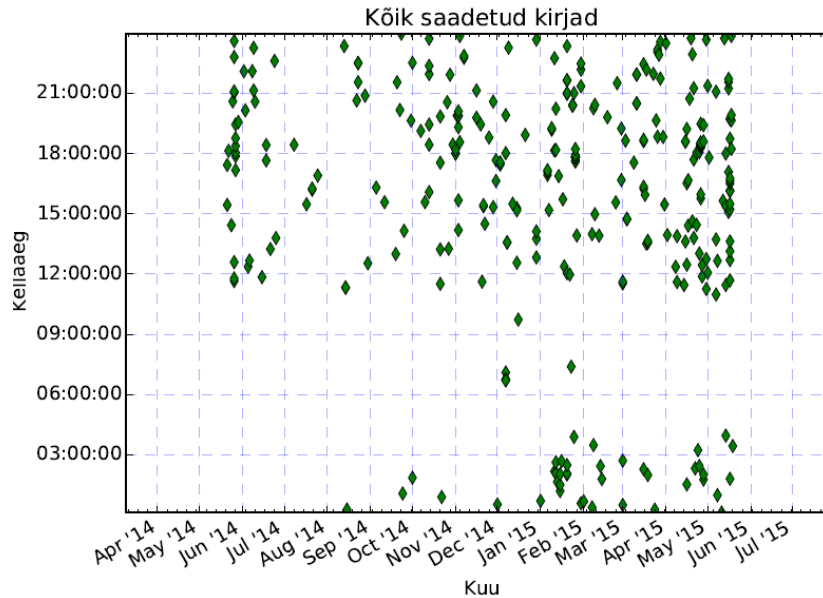
### 5.1 EDA tulemused ja analüüs

Programmi testimiseks kasutati spetsiaalselt loodud Gmaili kasutajat *vladimir.lopp@gmail.com*. Rakenduse tulemuste täielikuks demonstreerimiseks on postkast liiga väike. Seega lõpptestimiseks kasutati autori isiklikku meilikasti, kus oli kolmeaastase kasutusega kogunenud piisavalt andmeid.

Lõpptestimisel valiti postkasti analüüsimiseks järgmised otsinguparameetrid: postkasti ja saadetud kirjade kaustad, viimase aasta kirjad, võttesõna ei valitud. Statistika parameetrites märgistati kõik märkeruudud. Kirjade importimise peale läks aega ainult 8 minutit ja 6 sekundit. Võttes arvesse kirjade hulga (umbes 990) on see küllaltki kiire. Graafikud olid koostatud korrektselt ja kandsid endaga kaasas palju informatsiooni.



Joonis 5.1: Kõik postkasti kirjad viimase aasta lõikes

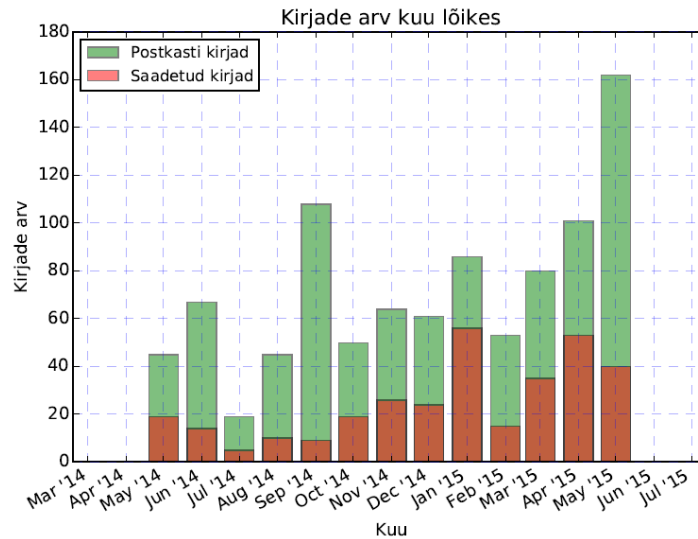


Joonis 5.2: Kõik saadetud kirjad viimase aasta lõikes

Joonis 5.1 näitab, kui palju kirju on saadud teatud kellaajal viimase aasta lõikes. Joonis 5.2 näitab, kui palju kirju on saadetud. Esimese pilguga on näha, et autorile on kirju saadetud palju rohkem, kui ta ise saatnud on. Samuti on mõlemast graafikust näha, et suveperiood oli väheaktiivsem. Koolisemestri lõpus (juuni ja jaanuar) on kirju nii saadetud, kui ka saabunud palju rohkem, kui teistel aegadel. Sama jutt käib ka semestri alguse kohta septembris. Lisaks on selgelt näha, et suurem osa kirju liigub alates kella 8 hommikul, kuni kella 1 öösel. Peale selle on mõnes kohas märgatavad sarnased sümbolite grupid mõlemas graafikus, mis asuvad sama koha peal. See tähendab, et siis on toimunud tihe kirjavahetus, kui umbes samal ajahetkel on autor saatnud kirju ja talle koheselt vastati.

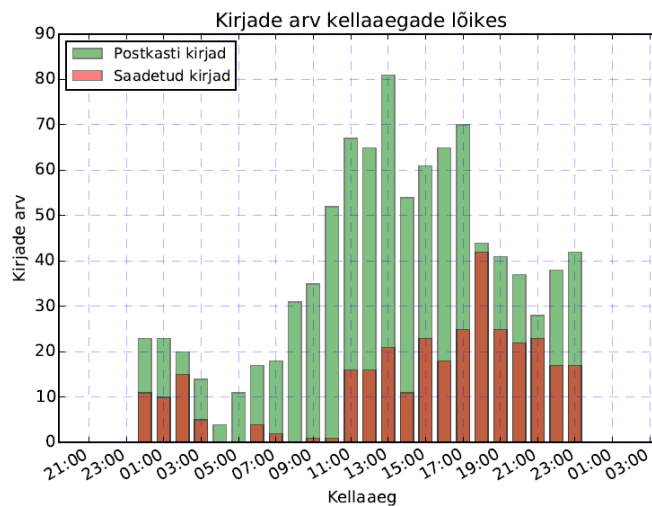
Tavaliselt töö autor kustutab tema meilikasti tulnud Facebooki teated, aga testimise jaoks on ta mai keskel tulnud teated jätnud kustutamata, mida on joonisel 5.1 ka selgelt näha.





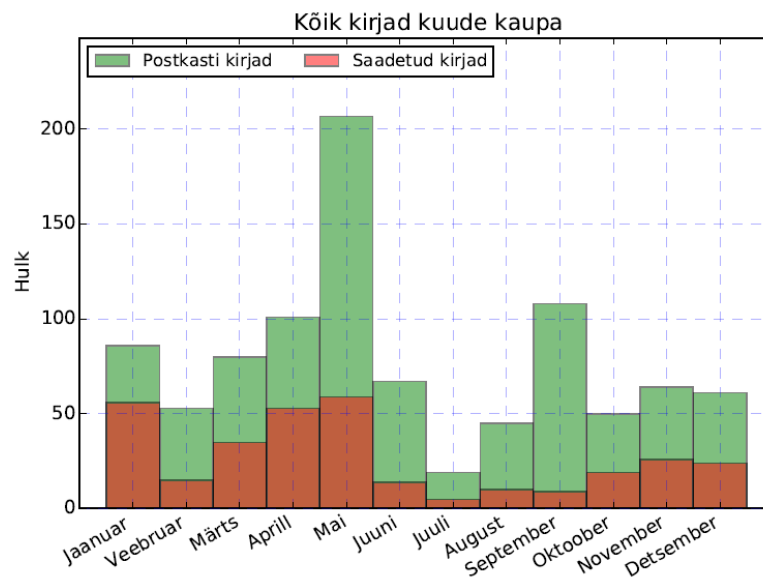
Joonis 5.3: Kirjade arv kuude lõikes

Joonis 5.3 näitab viimasel aastal saadetud ja saadud kirju kuude järgi. Siit on ka näha, et suveperiood on väheaktiivsem ja, et semestri lõpud on aktiivsemad. Samuti on märgatav mai pikk tulp, mille kuu Facebooki teated pole autor kustutanud. Samasugust graafikut väljastas programm ka ajadiagrammina.

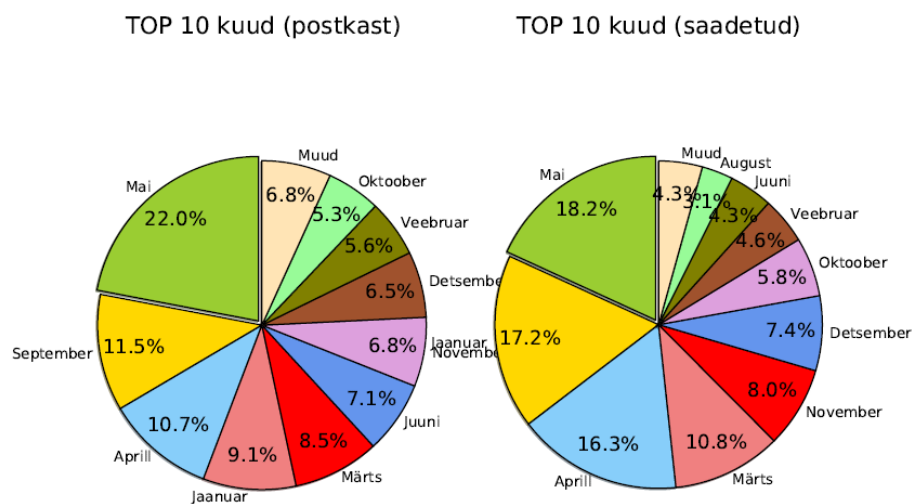


Joonis 5.4: Kirjade arv kellaaegade lõikes

Joonis 5.4 näitab, mis kellaaegadel on autor saanud ja saatnud oma kirju. Eriti eristuvad öised kellaajad, mis on üpris loogiline, kuna siis inimesed tavaliselt magavad ja aktiivset kirjavahetust ei toimu. Graafikult saab välja lugeda autori rutiini, et ta saadab harva kirju peale kella kahte öösel ning kohe hakkab oma Gmaili aktiivselt kasutama kella 11-st päeval. Seega autor magab tavaliselt antud kellaaegade vahel. Samasugust graafikut väljastas programm ka ajadiagrammina.

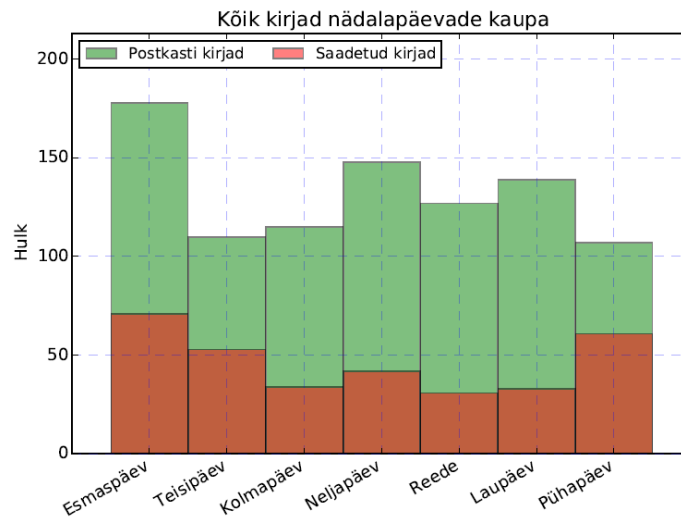


Joonis 5.5: Kõik kirjad kuude kaupa

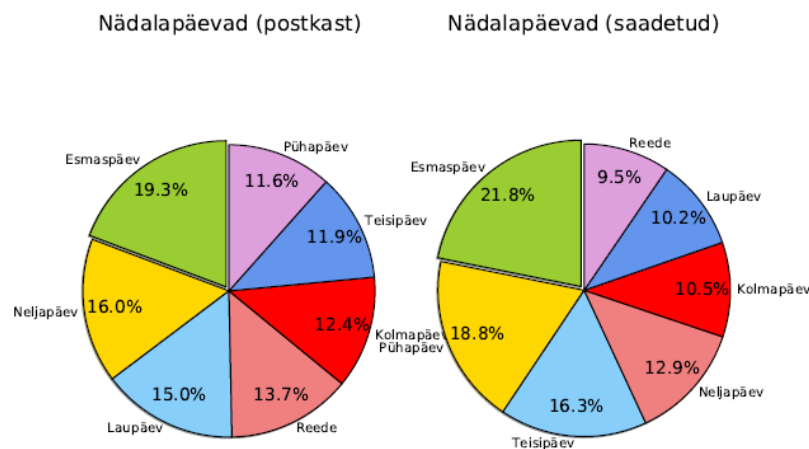


Joonis 5.6: Top kuud saadetud ja saadud kirjade kohta

Joonis 5.5 näitab saadetud ja saadud kirjade arvu kuude kaupa, vaatamata valitud ajavahemikule. Kuna antud testimise korral oli valitud viimase aasta ajavahemik, siis see graafik ei erine joonisel 5.3 esitatud graafikust. Joonisel 5.6 on samad andmed kujutatud pirukadiagrammidena.

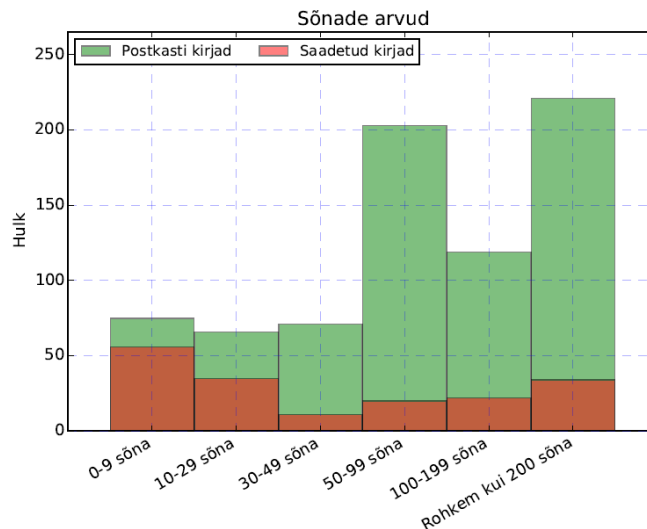


Joonis 5.7: Kõik kirjad nädalapäevade kaupa



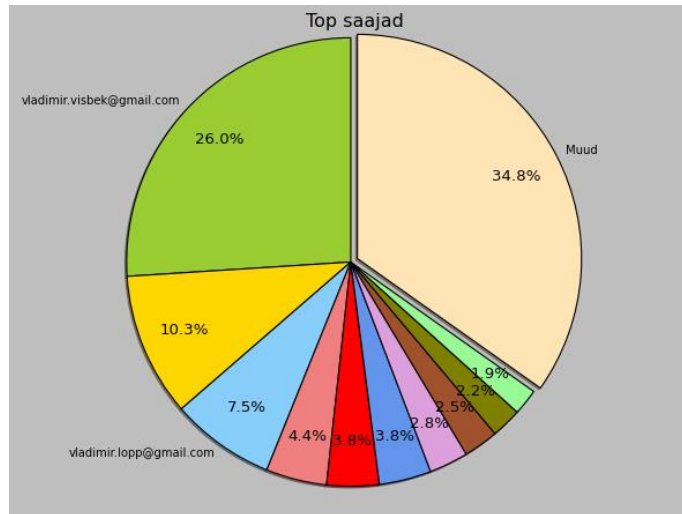
Joonis 5.8: Top nädalapäevad saadetud ja saadud kirjade kohta

Joonis 5.7 näitab saadetud ja saadud kirjade arvu nädalapäevade kaupa. Eelkõige paistab silma see, et uue nädala lähenemisega ja algusega saadetakse ja võetakse vastu rohkem kirju. Samas nädala sees on suhteliselt stabiilne aktiivsus. Joonisel 5.8 on samad andmed pirukadiagrammidena.



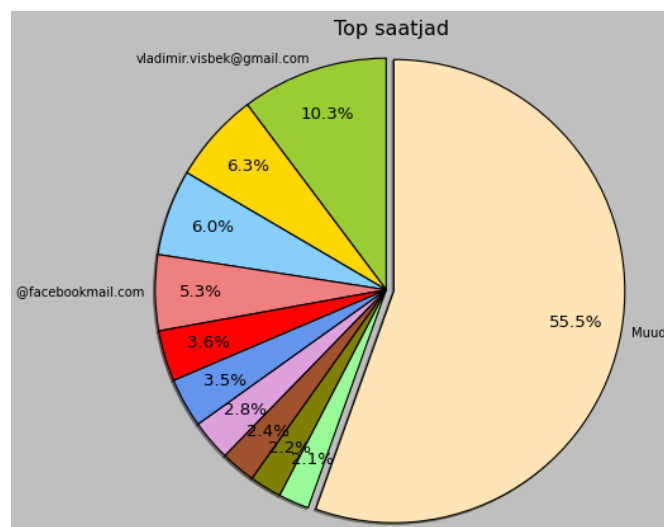
Joonis 5.9: Sõnade arvude tulpdiaagramm

Joonisel 5.9 on näha, kui pikad kirjad on kasutaja saanud ja saatnud. Kõige rohkem saadetakse kirju pikkusega kuni 9 sõna ning võetakse vastu sõnumeid, kus on üle 200 sõna. Arvatavasti pikemad kirjad on igasugused teated või reklaamid, kuna isiklikud kirjad on küllaltki harva nii pikad.



Joonis 5.10: Top 10 meilikasutajat, kellele autor kirju saatis

Joonis 5.10 näitab, mis kasutajatele ja kui palju autor kirju saadab. Kõige populaarsem on tema enda meiliaadress, mis tähendab, et ta saadab endale kõige rohkem kirju. Samuti on toimunud suurem meilivahetus Gmaili kasutajaga, *vladimir.lopp@gmail.com*, mida autor tekitas EDA programmi testimiseks ja debugimiseks. Mitmed teised meiliaadressid on ära kustutatud, kuna tegemist on isikliku postkastiga.



Joonis 5.11: Top 10 meilikasutajat, kes on autorile kirju saatnud

Vaatleme veel joonist 5.11, mis on analoogne joonisega 5.10. Antud juhul saab näha, kes on autorile kõige rohkem kirju saatnud. Siit on jällegi märgatav, et ta on ise endale kõige rohkem kirju saatnud. Samuti on olemas Facebooki teadete meiliaadress, mis oma mõne päeva teadetega on topi jõudnud. Ülejäänud kasutajad on jällegi peidetud.

## **5.2 Võrdlus teiste meiliklientidega**

Võrreldes teiste laiendustega ja programmidega, kus saab luua postkasti andmete põhjal graafikuid on EDA programmil nii eeliseid, kui ka puuduseid.

### **EDA eelised Immersiooni ees:**

- Rohkem tagasisidet ja graafikuid – top nädalapäevad ja kuud, kirjade arv kellaaja lõikes jne

### **EDA eelised Gmail Meter'i ees:**

- Otsingu ajavahemik – saab valida ajavahemiku, mille seast kirju otsitakse
- Kiirem – ei otsi enim kasutatavat sõna, seega töötlemist on vähem

### **EDA eelised mõlema ees:**

- Valikuvõimalus – saab valida parameetrid ja kriteeriumid, mille järgi kirju analüüsitakse
- Kaustade valik – postkasti ja saadetud kirjade kaustad saab eraldi valida
- Otsing võtmesõna(de) järgi – mitme võtmesõna korral saab sisestada loogilise avaldisena
- Kirjade sisu – võimalus vaadata leitud kirjade sisu
- Tulemus PDFis – võimalus tagasisidena saadud graafikud salvestada PDF failina

### **Immersion eelised:**

- Disain huvitavam – tulemuste esitamine visuaalse võrgustikuna on huvitavam
- Kiirem – Immersion ei analüüsi sisu, seega on vähem andmeid, mida töödelda

- Turvalisem – kasutab ainult kirjade meta-andmeid, seega kirjade sisu on rohkem kaitstud

#### **Gmail Meter'i eelised:**

- Saadab ise statistikat – saadab kasutaja meilile iga kuu alguses statistikat eelmise kuu kohta

### **5.3 Järeldused**

Antud töö raames koostatud rakendus EDA näitab, et programmeerimiskeelega Python on võimalik nullist luua meilirakendus, millega saab antud töö eesmärgid täies mahus realiseerida, st saab luua erinevat liiki graafikuid, et kasutaja saaks „pildi“ oma meilivahetusest.

## Kokkuvõte

Tänapäeval on meili teel suhtlemine muutunud väga populaarseks ning selleks on loodud mitmed meilikliendid, arvutiprogrammid ja veebimeilid. Töös uuritud kõige kasutatavamal meilikliendil Gmailil puudub võimalus postkasti andmete põhjalikumaks analüüsimiseks ja selle põhjal graafikute ja statistika koostamiseks. Seda on võimalik teha ainult kasutades selleks spetsiaalselt programmeeritud Gmaili laiendusi ja lisafunktsioone, nagu näiteks Immersion ja Gmail Meter.

Seega valmis antud töö raames Pythoni keeles programmeeritud meilirakendus, millega on võimalik sisse logida oma Gmaili kasutajaga, valida sobivad otsingukriteeriumid ja analüüsimise parameetrid, allalaadida nende põhjal oma postkastist kirju, kuvada need ekraanile ning luua nende kohta seitse huvitavat graafikut. Kõik nimetatud funktsionaalsused töötasid ka lõpptestimise ajal.

Testimise ja teiste meilirakendustega võrdlemise järel jõuti järeldusele, et EDA programmiga on võimalik analüüsida kasutaja Gmaili postkasti andmed ning näidata tulemust visualiseeritud graafikutena.



## **Summary**

### **Email box data visualization application**

Vladimir Visbek

Communicating via email has become very popular nowadays and there are several mail clients and programs for just that. The most popular one, Gmail, does not have the option to analyze your inbox data and show graphs and statistics based on your mail. It is only possible if you use certain Gmail extensions like Immersion and Gmail Meter.

As a result of this thesis, a mail application written in Python was created, which allows you to log in with your Gmail account, specify search and analyzation parameters, download mail and make seven interesting graphs based on those parameters. All of the abovementioned functionalities were working during final testing.

After testing and comparing it to other mail applications, a conclusion was reached that it is possible to analyze the data of the user's Gmail inbox and show the results visually, as graphs, using the EDA program.

## Tänuavaldused

Eelkõige soovib autor tänada oma juhendajat, Tõnu Tamme, lõputöö juhendamise ja mõistva suhtumise eest, kuna ilma tema nõu ja suunamiseta ei oleks käesolev töö valminud.

Lisaks avaldab autor suured tänud, töö valmimisel saadud abi ja nõuannete eest, järgmistele inimestele:

1. Siim Arras
2. Agris Hüüp
3. Olesja Rešnjak
4. Viktoria Visbek
5. Eduard Visbek

## Viited

- [1] Wikipedia the free encyclopedia. Carbon copy. [Online].  
[http://en.wikipedia.org/wiki/Carbon\\_copy](http://en.wikipedia.org/wiki/Carbon_copy) [Viimati vaadatud 22.05.2015]
- [2] Pobox.com. How to read email timestamps (date and time entries). [Online].  
<http://www.pobox.com/helpspot/index.php?pg=kb.page&id=256> [Viimati vaadatud 22.05.2015]
- [3] Wikipedia the free encyclopedia. Network socket. [Online].  
[http://en.wikipedia.org/wiki/Network\\_socket](http://en.wikipedia.org/wiki/Network_socket) [Viimati vaadatud 22.05.2015]
- [4] Wikipedia the free encyclopedia. Graphical user interface. [Online].  
[http://en.wikipedia.org/wiki/Graphical\\_user\\_interface](http://en.wikipedia.org/wiki/Graphical_user_interface) [Viimati vaadatud 22.05.2015]
- [5] Wikipedia the free encyclopedia. Email Client. [Online].  
[https://en.wikipedia.org/wiki/Email\\_client](https://en.wikipedia.org/wiki/Email_client) [Viimati vaadatud 22.05.2015]
- [6] Litmus Email Analytics. Email client market share. [Online].  
<http://emailclientmarketshare.com/> [Viimati vaadatud 22.05.2015]
- [7] Wikipedia the free encyclopedia. Gmail. [Online]. <http://en.wikipedia.org/wiki/Gmail>  
[Viimati vaadatud 22.05.2015]
- [8] Wikipedia the free encyclopedia. Conversation threading. [Online].  
[http://en.wikipedia.org/wiki/Conversation\\_threading](http://en.wikipedia.org/wiki/Conversation_threading) [Viimati vaadatud 22.05.2015]
- [9] Gmail. Gmail Help. [Online]. <https://support.google.com/mail/answer/3055016?hl=en>  
[Viimati vaadatud 22.05.2015]

- [10] Liz Jacobs. (2013, Oktoober) Immersion: A behind-the-scenes look at a tool that maps your life based on every email you've ever sent. [Online]. <http://blog.ted.com/immersion-a-behind-the-scenes-look-at-a-tool-that-maps-your-life-based-on-every-email-youve-ever-sent/> [Viimati vaadatud 22.05.2015]
- [11] MIT lab. (2013) Immersion. [Online]. <https://immersion.media.mit.edu/> [Viimati vaadatud 22.05.2015]
- [12] Gmail Meter. (2015) Advanced Email Analytics & Statistics. [Online]. <http://www.gmailmeter.com/> [Viimati vaadatud 22.05.2015]
- [13] Saurabh Gupta. (2012, Aprill) Know Your Gmail Stats using Gmail Meter. [Online]. <http://gmailblog.blogspot.com/2012/04/know-your-gmail-stats-using-gmail-meter.html> [Viimati vaadatud 22.05.2015]
- [14] Madis Raud, "Comparison of email browsers and their search capabilities," University of Tartu, Bachelor's Thesis 2012.
- [15] Python Documentation. (2015) 20.10. imaplib — IMAP4 protocol client. [Online]. <https://docs.python.org/2/library/imaplib.html> [Viimati vaadatud 22.05.2015]
- [16] Robert Iwancz. (2013, Juuli) Using Python to get email from a Gmail account via IMAP: The Least You Need To Know. [Online]. <http://www.voidynullness.net/blog/2013/07/25/gmail-email-with-python-via-imap/> [Viimati vaadatud 22.05.2015]
- [17] PyMOTW. (2014, August) imaplib - IMAP4 client library. [Online]. <http://pymotw.com/2/imaplib/> [Viimati vaadatud 22.05.2015]
- [18] Stackoverflow. (2011, Aprill) SEARCH BEFORE/AFTER with Python's imaplib. [Online]. <http://stackoverflow.com/questions/5621341/search-before-after-with-pythons-imaplib> [Viimati vaadatud 22.05.2015]

- [19] Python Documentation. (2015) 18.1.1. email.message: Representing an email message. [Online].  
[https://docs.python.org/2/library/email.message.html#email.message.Message.is\\_multipart](https://docs.python.org/2/library/email.message.html#email.message.Message.is_multipart)  
[Viimati vaadatud 22.05.2015]
- [20] Stackoverflow. (2009, September) How can I get an email message's text content using python? [Online]. <http://stackoverflow.com/questions/1463074/how-can-i-get-an-email-messages-text-content-using-python> [Viimati vaadatud 22.05.2015]
- [21] Laura Creighton. (2015, Jaanuar) GUI Programming in Python. [Online].  
<https://wiki.python.org/moin/GuiProgramming> [Viimati vaadatud 22.05.2015]
- [22] Dale Athanasias. (2014, Juuni) TkInter. [Online]. <https://wiki.python.org/moin/TkInter>  
[Viimati vaadatud 22.05.2015]
- [23] Matplotlib. (2015, Veebruar) Usage. [Online]. [http://matplotlib.org/faq/usage\\_faq.html](http://matplotlib.org/faq/usage_faq.html)  
[Viimati vaadatud 22.05.2015]

## **Lisad**

Töö raames valminud meilirakendusele EDA saab ligi järgmise lingi kaudu:

[https://www.dropbox.com/s/6kryio9wfsytcu/EDA\\_visbek.zip?dl=0](https://www.dropbox.com/s/6kryio9wfsytcu/EDA_visbek.zip?dl=0)

# **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Vladimir Visbek

1. Annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

## **"RAKENDUS MEILIPOSTKASTI ANDMETE VISUALISEERIMISEKS"**

mille juhendaja on Tõnu Tamme

- (a) reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - (b) üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi Dspace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile;
  3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **22.05.2015**